Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE DIRECT

Journal of Multivariate Analysis 87 (2003) 24–45

Journal of Multivariate Analysis

http://www.elsevier.com/locate/jmva

Identification of graphical models for nonignorable nonresponse of binary outcomes in longitudinal studies

Wen-Qing Ma,[a] Zhi Geng,[b,*] and Yong-Hua Hu[a]

[a] *School of Public Health, Peking University, Beijing 100083, China*
[b] *Department of Probability and Statistics, School of Mathematical Sciences, Peking University, Beijing 100871, China*

Received 20 July 2000

Abstract

In this paper, we use directed acyclic graphs (DAGs) with temporal structure to describe models of nonignorable nonresponse mechanisms for binary outcomes in longitudinal studies, and we discuss identification of these models under an assumption that the sequence of variables has the first-order Markov dependence, that is, the future variables are independent of the past variables conditional on the present variables. We give a stepwise approach for checking identifiability of DAG models. For an unidentifiable model, we propose adding completely observed variables such that this model becomes identifiable.

AMS 2000 subject classifications: 62H17; 62H12; 62H20

Keywords: Binary outcome; Directed acyclic graph; Identifiability; Longitudinal data; Nonignorable nonresponse

1. Introduction

Nonresponse in a longitudinal study often arises due to intermittent visits or drop-out. Various models for longitudinal categorical data subject to nonignorable nonresponse have been discussed by many investigators, see [1,4,8–10]. The models for nonresponse mechanisms can be classified into two types, ignorable and nonignorable. For a likelihood-based inference, it is not necessary to introduce a

*Corresponding author.
E-mail address: zgeng@math.pku.edu.cn (Z. Geng).

0047-259X/03/$ - see front matter
doi:10.1016/S0047-259X(03)00043-5

model for an ignorable nonresponse mechanism, but a model is needed to describe a nonignorable nonresponse mechanism. Diggle and Kenward [3] proposed a logistic regression model for nonignorable drop-out mechanism. Rotnitzky et al. [13] proposed a class of estimators with responses weighted by the reciprocals of the probabilities. Identification of parameters is a fundamental problem for nonresponse mechanisms. Rothenberg [12] described definitions of local and global identification and proved that local identifiability is equivalent to nonsingularity of the information matrix. Fitzmaurice et al. [5] presented some simple procedures for checking the 'global' identifiability in a selected range of parameter values. Glonek [6] presented necessary and sufficient conditions of global identifiability for simple nonignorable nonresponse models with one or two binary responses. Fay [4] proposed a directed acyclic graph (DAG) for a nonresponse mechanism with two response variables, and Little [8] discussed DAGs for longitudinal studies with two time points. In this paper, we use DAGs with temporal structure to describe models of nonignorable nonresponse mechanisms for binary outcomes in longitudinal studies. We discuss identification of these models under the assumption that the sequence of variables has the first-order Markov dependence, that is, the future variables are independent of the past variables conditional on the present variables. We describe conditions for identifiability, and present a stepwise approach for checking identifiability of models. This approach asserts identifiability qualitatively on the basis of structures of DAGs, and thus it can be applied to a study design. The approach cannot assert that a model is not identifiable since most conditions presented in this paper are sufficient and not necessary for identifiability. For an unidentifiable model, we propose to add a completely observed variable to the model; with its addition, the model may become identifiable.

In Section 2, we describe the DAG models for nonresponse mechanisms for binary outcomes. Section 3 discusses the conditions of identifiability. Section 4 presents a stepwise approach for checking identifiability. In Section 5, we propose to add a completely observed variable to an unidentifiable model so that the model may become identifiable. All the proofs are given in the appendix.

2. Directed acyclic graphs for nonresponse mechanisms

Suppose that individuals are observed repeatedly T times in a longitudinal study. Consider a binary outcome Y_t with a value 0 or 1 at time point t, $t = 1, \ldots, T$. For each outcome Y_t, we introduce a response indicator R_t with value 1 if the outcome Y_t is obtained and 0 otherwise. Assume that the sequence $(Y_1, R_1, \ldots, Y_T, R_T)$ of outcomes and response indicators has the first-order Markov dependence, that is, (Y_{t+1}, R_{t+1}) is conditionally independent of $(Y_1, R_1, \ldots, Y_{t-1}, R_{t-1})$ given (Y_t, R_t). This is denoted as

$$(Y_{t+1}, R_{t+1}) \perp\!\!\!\perp (Y_1, R_1, \ldots, Y_{t-1}, R_{t-1}) | (Y_t, R_t).$$

This assumption states that the outcome and missingness history from time 1 to $t-1$ is not predictive of the outcome and missingness at time $t+1$ after adjusting for the

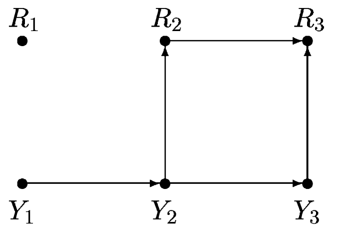

Fig. 1. A DAG model G with three time points.

outcome and missingness at time t. We use a DAG model to describe such a nonresponse mechanism. A DAG is denoted by $G = (V, E)$ where $V = \{Y_1, R_1, \ldots, Y_T, R_T\}$ is the set of nodes, $E = \{\langle v_1, v_2 \rangle | v_1 \neq v_2, v_1 \in V \text{ and } v_2 \in V\}$ is the set of arrows and $< v_1, v_2 >$ denotes an arrow from v_1 to v_2. A subgraph from time point s to t is denoted by $G_{s,t} = (V_{s,t}, E_{s,t})$ where $V_{s,t} = \{Y_s, R_s, Y_{s+1}, R_{s+1}, \ldots, Y_t, R_t\}$ and $E_{s,t} = \{\langle v_1, v_2 \rangle | \langle v_1, v_2 \rangle \in E, v_1 \in V_{s,t}$ and $v_2 \in V_{s,t}\} = E \cap (V_{s,t} \times V_{s,t})$ (i.e. the arrows of G appear among $V_{s,t}$). The joint probability of $(Y_1, R_1, \ldots, Y_T, R_T)$ according to this DAG is described by

$$P(y_1, r_1, \ldots, y_T, r_T) = \prod_{x \in V} P(x|pa_x),$$

where pa_x denotes the set of parents of x, see [2,7,11]. For a DAG G, let $M(G)$ denote the set of all joint probabilities that can be described according to G, and let $M(G)_{s,t} = \{P(y_s, r_s, \ldots, y_t, r_t) | P(y_1, r_1, \ldots, y_T, r_T) \in M(G)\}$ denote the set of marginal probabilities.

Fig. 1 shows an example of a DAG model with three time points. The joint probability of $(Y_1, R_1, Y_2, R_2, Y_3, R_3)$ according to this DAG G can be expressed as follows:

$$\begin{aligned} P(y_1, r_1, y_2, r_2, y_3, r_3) &= \prod_{x \in V} P(x|pa_x) \\ &= P(y_1)P(r_1)P(y_2|y_1)P(r_2|y_2)P(y_3|y_2)P(r_3|y_3, r_2). \end{aligned}$$

In this DAG model, (Y_1, R_1) and (Y_3, R_3) are conditionally independent given (Y_2, R_2), Y_1 is missing at random, the response indicator R_2 depends on value of Y_2, and the response indicator R_3 depends on both values of Y_3 and R_2.

3. Identifiability of DAG models

For the DAG models introduced in the previous section, we present a stepwise approach for checking identification. We check the identifiability of a model separately at each time point, in the chronological order. A DAG model may be unidentifiable at the beginning but become identifiable at a later time point. For example, it will be shown at the end of Section 4 that the model in Fig. 2 is not

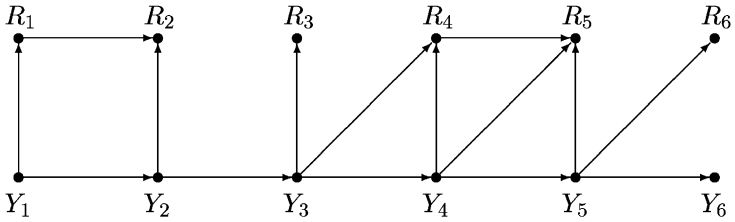

Fig. 2. A DAG model with 6 time points.

identifiable if the study stops at the time point 2, but it becomes identifiable if the study continues to time point 3. We introduce below definitions of period identifiability and conditional identifiability.

Let $D_{s,t}$ $(s \leqslant t)$ denote the data obtained from the time point s to t, that is, the observed frequencies of $(r_s, r_{s+1}, \dots, r_t; y_q$ for $r_q = 1$ and $q = s, \dots, t)$. For example, $D_{t,t+1}$ denotes the observed frequencies of $(R_t = 1, R_{t+1} = 1, y_t, y_{t+1})$, $(R_t = 1, R_{t+1} = 0, y_t)$, $(R_t = 0, R_{t+1} = 1, y_{t+1})$ and $(R_t = 0, R_{t+1} = 0)$ for y_t and $y_{t+1} = 0$ and 1.

Given a family of distributions, say $M(G)$, we say that every distribution P in $M(G)$ is identifiable if, given a distribution P^* of the observed data $D_{1,T}$, there is a unique distribution P in $M(G)$ that has the same marginal distribution for the observed data as P^*. Now we define identifiability of graphical models.

Definition 1. A subgraph $G_{s,t}$ $(s \leqslant t)$ is YR-identifiable if every distribution $P(y_s, r_s, y_{s+1}, r_{s+1}, \dots, y_t, r_t)$ in $M(G)_{s,t}$ is identifiable.

Definition 2. A subgraph $G_{s,t}$ $(s \leqslant t)$ is Y-identifiable if every distribution $P(y_s, y_{s+1}, \dots, y_t)$ in $M(G)_{s,t}$ is identifiable.

From the above definitions, the YR-identifiability implies the Y-identifiability, but the converse is not necessarily correct. We introduce below two concepts of conditional identifiability, that is, identifiability of a subgraph $G_{s,t}$ dependent on the structure of G at other time points. In our stepwise approach to checking identifiability, conditional identifiability means that identifiability cannot be asserted at the present time point but may be asserted at some later time points.

Definition 3. A subgraph $G_{s,t}$ $(s \leqslant t)$ is conditionally YR-identifiable if every distribution $P(y_s, r_s, y_{s+1}, r_{s+1}, \dots, y_t, r_t)$ in $M(G)_{s,t}$ is identifiable when at least one of the marginal distributions $P(y_i)$ for $s \leqslant i \leqslant t$ is identifiable.

Definition 4. A subgraph $G_{s,t}$ $(s \leqslant t)$ is conditionally Y-identifiable if every marginal distribution $P(y_s, y_{s+1}, \dots, y_t)$ in $M(G)_{s,t}$ is identifiable when at least one of the marginal distributions $P(y_i)$ for $s \leqslant i \leqslant t$ is identifiable.

Definition 5. A subgraph $G_{s,t}$ $(s \leqslant t)$ is YR-identifiable by $D_{s,t}$ if every distribution $P(y_s, r_s, y_{s+1}, r_{s+1}, \ldots, y_t, r_t)$ in $M(G)_{s,t}$ is identifiable by $D_{s,t}$.

Identifiability in Definitions 1–4 means that subgraphs can be identified by using all observed data, but identifiability by $D_{s,t}$ in Definition 5 means that they can be identified by using only a part of observed data, $D_{s,t}$. Thus the YR-identifiability of $G_{s,t}$ by $D_{s,t}$ implies the YR-identifiability of $G_{s,t}$, but the converse is not necessarily correct. Similarly we can define Y-, conditional YR- and conditional Y-identifiability by $D_{s,t}$.

For the DAG model in Fig. 2, we shall show the following identifiability at the end of Section 4. The subgraphs $G_{1,3}$ is YR-identifiable, $G_{4,5}$ is Y-identifiable and $G_{5,6}$ is YR-identifiable. If the study stops at time point 2, then $G_{1,2}$ is conditionally YR-identifiable, and not YR-identifiable by $D_{1,2}$. But $G_{1,2}$ becomes YR-identifiable by $D_{1,3}$ if the study continues to time point 3. Similarly, the subgraph $G_{4,5}$ is Y-identifiable by $D_{4,6}$, but not Y-identifiable by $D_{4,5}$.

We first consider identifiability of a subgraph at the first time point, $t = 1$. If a subgraph $G_{1,1}$ does not have the arrow from Y_1 to R_1 as shown in Fig. 3(a), then Y_1 is independent of R_1, denoted by $Y_1 \perp\!\!\!\perp R_1$; that is, Y_1 is missing completely at random. We have that $P(y_1, r_1) = P(y_1 | R_1 = 1)P(r_1)$, and thus the subgraph $G_{1,1}$ is YR-identifiable by $D_{1,1}$. If a subgraph $G_{1,1}$ has the arrow from Y_1 to R_1 as shown in Fig. 3(b), then this model is saturated for the distribution of Y_1 and R_1. In this case, a distribution of the observed data $D_{1,1}$ cannot identify these distributions of Y_1 and R_1 which have the same values of $P(R_1 = 0) = P(Y_1 = 1, R_1 = 0) + P(Y_1 = 0, R_1 = 0)$ but different values of $P(y_1, R_1 = 0)$. Thus the subgraph (b) is not YR-identifiable by $D_{1,1}$.

In this case, we cannot assert the identifiability without checking the later time points. In Section 5, we shall propose adding a binary variable X to the model of Fig. 3(b) such that the subgraph becomes identifiable without necessity of checking the late time points, as shown in Fig. 3(c) where the undirected edge between X and Y_1 denotes an arrow from X to Y_1 or from Y_1 to X.

Next, we discuss identifiability of a subgraph $G_{t,t+1}$. There are two ways for specifying the joint distribution of outcomes and response indicators. One is

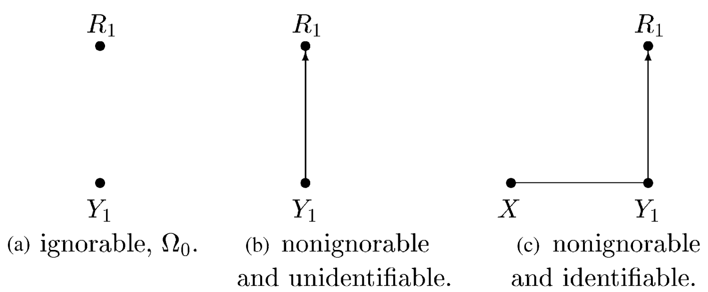

Fig. 3. Subgraphs $G_{1,1}$. (a) ignorable, Ω_0. (b) nonignorable and unidentifiable. (c) nonignorable and identifiable.

selection model and the other is pattern-mixture model. The selection model $P(y_1, \dots, y_T, r_1, \dots, r_T) = P(y_1, \dots, y_T)P(r_1, \dots, r_T|y_1, \dots, y_T)$ is used here and it means that dependence between an indicator and an outcome is described by using an arrow from the outcome to the indicator; there are no arrows from indicator to outcome. Thus there are six possible arrows among the four nodes Y_t, Y_{t+1}, R_t and R_{t+1} in the subgraph $G_{t,t+1}$. Assume that the response indicator R_t does not depend on the value of Y_{t+1} conditional on Y_t, that is, the present missingness does not depend on the future outcomes given the present outcome. Thus there is no arrow from Y_{t+1} to R_t in $G_{t,t+1}$. This assumption can decrease half the number of subgraph types. Without this assumption, identifiability of the other half of subgraph types can be derived in the same way as below. We always draw an arrow from Y_t to Y_{t+1} for any t, but allow $P(y_{t+1} \mid y_t) = P(y_{t+1})$. Then there are 16 possible graphs $G_{t,t+1}$ in total, as shown in Fig. 4. These graphs are classified into five sets, Ω_i for $i = 1, \dots, 5$. Graph (a) in Fig. 3 forms the set Ω_0. It will be shown below that the subgraphs in each set have the same condition for identifiability.

Define

$$\theta_{j|h}^{(t)} = P(Y_{t+1} = j, R_{t+1} = 1 \mid Y_t = h, R_t = 1), \quad \lambda_{hi}^{(t)} = P(Y_t = h, R_t = i),$$

$$p_{j|h}^{(t)} = P(Y_{t+1} = j \mid Y_t = h) \quad \text{and} \quad \rho_{hij}^{(t)} = P(R_{t+1} = 1 \mid Y_t = h, R_t = i, Y_{t+1} = j),$$

for all h, i and j. The identity $\rho_{h0j}^{(t)} = 0$ for a time point t means that nonresponse at time t implies nonresponse at time $t + 1$. Drop-out is a special case in which $\rho_{h0j}^{(t)} = 0$ for all t, h and j. In the remainder of this section, we consider only identifiability of subgraphs $G_{t,t+1}$ and so omit the superscript (t) of all parameters to simplify the notation. We can determine $\theta_{j|h}$ from $D_{t,t+1}$ directly. The joint distribution of $(Y_t$, R_t, Y_{t+1}, $R_{t+1})$ can be written as

$$P(Y_t = h, R_t = i, Y_{t+1} = j, R_{t+1} = k) = \lambda_{hi} p_{j|h} \rho_{hij}^k (1 - \rho_{hij})^{1-k}.$$

Thus YR-identifiability of $G_{t,t+1}$ means that all parameters λ_{hi}, $p_{j|h}$ and ρ_{hij} are identifiable. Assume that $P(Y_t = h, R_t = 1, Y_{t+1} = j, R_{t+1} = k) > 0$ for all h, j and k.

Lemma 1. *Parameters λ_{hi} and $P(y_t)$ are identifiable if $P(y_{t+1})$ and $p_{j|h}$ are identifiable and $Y_t \not\perp\!\!\!\perp Y_{t+1}$.*

In general, a typical proof of the identifiability for a given subgraph first determines which of variables the parameter ρ_{hij} depends on (e.g. for subgraph (c) of Fig. 4, it depends on Y_{t+1} only, that is, $\rho_{hij} = \rho_j$), then proves that ρ_{hij} can be uniquely determined by $\theta_{j|h}$, and finally shows that λ_{hi} and $p_{j|h}$ can be identified by $\theta_{j|h}$ and ρ_{hij}. Below we explore the identifiability separately for each set Ω_i. We show below that all subgraphs $G_{t,t+1}$ in Ω_1 are YR-identifiable by $D_{t,t+1}$ when $Y_{t+1} \not\perp\!\!\!\perp Y_t$.

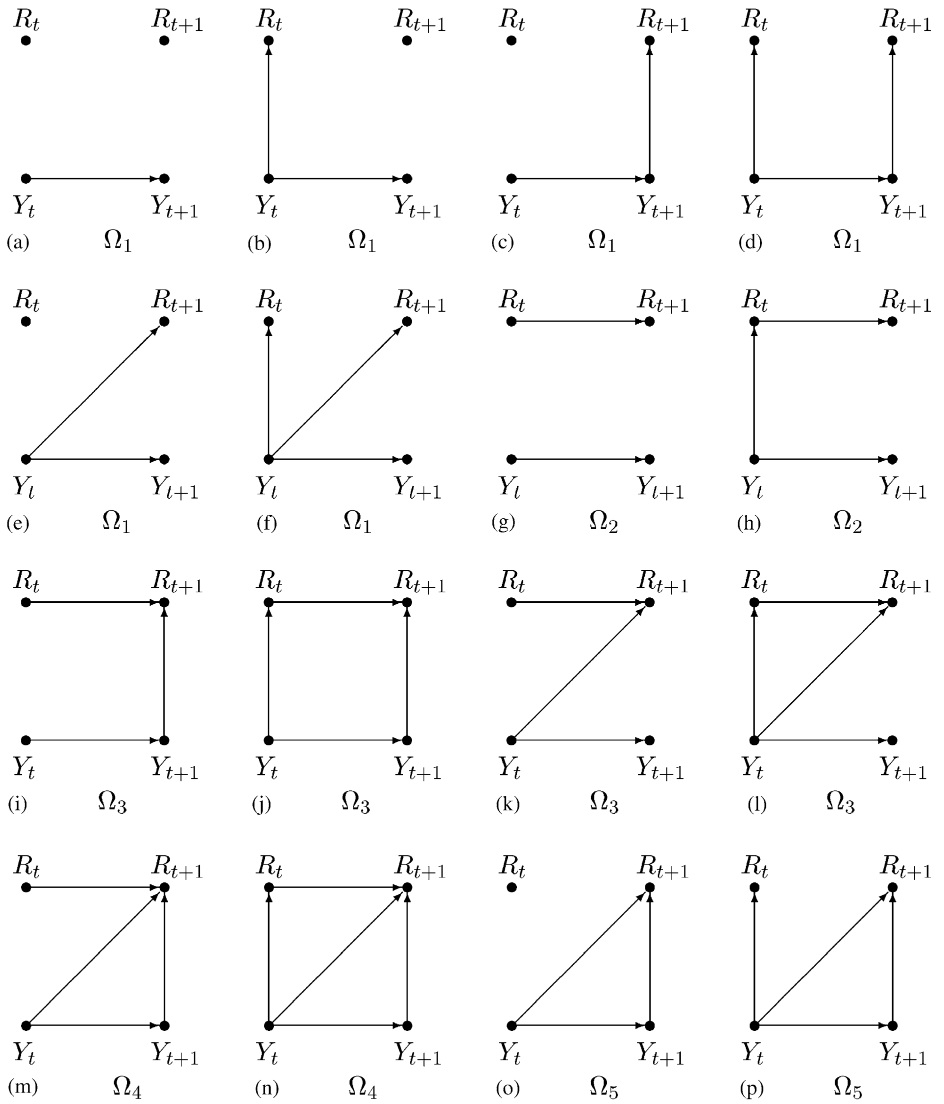

Fig. 4. Sixteen possible $G_{t,t+1}$. (a) Ω_1; (b) Ω_1; (c) Ω_1; (d) Ω_1; (e) Ω_1; (f) Ω_1; (g) Ω_2; (h) Ω_2; (i) Ω_3; (j) Ω_3; (k) Ω_3; (l) Ω_3; (m) Ω_4; (n) Ω_4; (o) Ω_5; (p) Ω_5.

Theorem 1. *For subgraphs* (a)–(f) *in the set* Ω_1, *we have that*

1. *subgraphs* (a) *and* (b) *are YR-identifiable by* $D_{t,t+1}$ *if and only if* $Y_{t+1} \not\perp\!\!\!\perp Y_t$ *or* $R_t \perp\!\!\!\perp Y_t$;
2. *subgraphs* (c) *and* (d) *are YR-identifiable by* $D_{t,t+1}$ *if and only if* $Y_{t+1} \not\perp\!\!\!\perp Y_t$; *and*
3. *subgraphs* (e) *and* (f) *are YR-identifiable by* $D_{t,t+1}$.

Since subgraphs (e) and (f) become (a) and (b), respectively, when $Y_t \perp\!\!\!\perp R_{t+1}$, we obtain, according to Theorem 1, that all subgraphs in Ω_1 are identifiable if $Y_{t+1} \not\perp\!\!\!\perp Y_t$.

We consider Ω_2 separately for the cases in which the missing process is drop-out or not.

Theorem 2. *For subgraphs* (g) *and* (h) *in the set* Ω_2, *we have that*

1. *when* $P(R_{t+1} = 1|R_t = 0) > 0$, *the subgraphs are YR-identifiable by* $D_{t,t+1}$ *if and only if* $Y_{t+1} \not\perp\!\!\!\perp Y_t$ *or* $R_t \perp\!\!\!\perp Y_t$;
2. *when* $P(R_{t+1} = 1|R_t = 0) = 0$, *the subgraphs are conditionally YR-identifiable by* $D_{t,t+1}$ *if* $Y_{t+1} \not\perp\!\!\!\perp Y_t$, *and they are YR-identifiable by* $D_{t,t+1}$ *if* $R_t \perp\!\!\!\perp Y_t$.

We consider the set Ω_3 of graphs (i)–(l) in Fig. 4, and we show below that all subgraphs $G_{t,t+1}$ in Ω_3 are conditionally YR-identifiable by $D_{t,t+1}$ if $Y_{t+1} \not\perp\!\!\!\perp Y_t$.

Theorem 3. *Subgraphs* (i)–(l) *in the set* Ω_3 *are conditionally YR-identifiable by* $D_{t,t+1}$ *if* $Y_{t+1} \not\perp\!\!\!\perp Y_t$.

We now consider the set Ω_4 of graphs (m) and (n) in Fig. 4. Because there are fewer observed frequencies in $D_{t,t+1}$ than the number of parameters for models (m) and (n), these two models are not YR-identifiable by $D_{t,t+1}$. We shall show below that they may be identified under certain conditions by using data obtained at more time points. We first give a lemma and then show that they are Y-identifiable by $D_{t,t+2}$ if $G_{t+1,t+2}$ belongs to Ω_1 or Ω_5.

Lemma 2. *Let* Z_{t+2} *be a binary variable at time point* $t+2$ (*such as* R_{t+2} *or* Y_{t+2}). *Suppose that* $R_{t+1} \perp\!\!\!\perp Z_{t+2} \mid Y_{t+1}$, $R_t \perp\!\!\!\perp Y_{t+1} \mid Y_t$ *and* $P(y_{t+1}, R_{t+1} = 1 \mid y_t, R_t = 1, Z_{t+2} = z)$ *is identifiable by* $D_{t,t+2}$ *for any* y_t, y_{t+1} *and* z. *Then both* $p_{j|h}$ *and* ρ_{h1j} *are identifiable by* $D_{t,t+2}$ *if* $Z_{t+2} \not\perp\!\!\!\perp Y_{t+1}$.

We show below a result for Ω_5 which is needed for discussing identification of Ω_4,

Theorem 4. *For subgraphs* (o) *and* (p) *in* Ω_5, $P(y_t, r_t)$ *is identifiable by* $D_{t,t+1}$ *if* $R_{t+1} \not\perp\!\!\!\perp Y_t$.

Theorem 5. *Subgraphs* (m) *and* (n) *in the set* Ω_4 *are Y-identifiable by* $D_{t,t+2}$ *if* (i) $Y_{t+1} \not\perp\!\!\!\perp Y_t$ *and* (ii) ($G_{t+1,t+2} \in \Omega_1$ *and* $Y_{t+2} \not\perp\!\!\!\perp Y_{t+1}$) *or* ($G_{t+1,t+2} \in \Omega_5$ *and* $R_{t+2} \not\perp\!\!\!\perp Y_{t+1}$).

Theorem 6. *Suppose that conditions* (i) *and* (ii) *of Theorem* 5 *hold. We have*

1. *subgraphs* (m) *and* (n) *in the set* Ω_4 *are YR-identifiable by* $D_{t,t+2}$ *if the missing process is drop-out at time points* t *and* $t+1$; *and*
2. *subgraphs* (m) *and* (n) *in the set* Ω_4 *are YR-identifiable by* $D_{t-1,t+2}$ *if* $G_{t-1,t}$ *is a subgraph in* Ω_1, Ω_2 *or* Ω_3, *and* $Y_{t-1} \not\perp\!\!\!\perp Y_t$.

Finally, we consider the set Ω_5 of graphs (o) and (p) in Fig. 4, and prove their identifiability by checking the structures of graphs at the next two time points, $G_{t+1,t+2}$.

Theorem 7. *Subgraphs* (o) *and* (p) *in* Ω_5 *are* YR*-identifiable by* $D_{t,t+2}$ *if* (i) $Y_{t+1} \not\perp\!\!\!\perp Y_t$ *and* (ii) $(G_{t+1,t+2} \in \Omega_1$ *and* $Y_{t+2} \not\perp\!\!\!\perp Y_{t+1})$ *or* $(G_{t+1,t+2} \in \Omega_5$ *and* $R_{t+2} \not\perp\!\!\!\perp Y_{t+1})$.

4. A stepwise approach for checking identifiability

We present a stepwise approach for checking identifiability of DAG models for nonignorable nonresponse mechanisms based on the results of Section 3. For a given DAG model, this approach checks the identifiability simply one-by-one time point. It does not consider the cases of perfect cancellation of association, that is, when some variables are not separated in the structure of a DAG but are coincidentally independent. For instance, it is possible that $R_{t+1} \perp\!\!\!\perp Y_t$ for the subgraphs in Ω_5. The perfect cancellation cannot be checked in a study design. The approach does not depend on quantitative information and thus it can be applied in both a study design and data analysis.

We assume that $Y_t \not\perp\!\!\!\perp Y_{t+1}$ for all t. Let $Id_i = YR$, Y, C and U for $i = s, s+1, \dots, t$ denote that the subgraph $G_{s-1,t}$ is YR-, Y-, conditional YR- and undecided identifiability, respectively. When $Id_s = Id_{s+1} = \cdots = Id_t = YR$ (or Y), the subgraph $G_{s-1,t}$ is YR- (or Y-) identifiable, and it means that every distribution for YR (or Y) in $M(G_{s-1,t})$ can be identified by using observed data $D_{1,T}$. $Id_s = Id_{s+1} = \cdots = Id_t = U$ means that the identifiability of $G_{s-1,t}$ cannot be asserted. $Id_s = Id_{s+1} = \cdots = Id_t = C$ denotes that the subgraph $G_{s-1,t}$ is conditionally YR-identifiable. For a conditionally identifiable subgraph $G_{s-1,t}$, it becomes identifiable if one of $P(y_i)$ for $i = s-1, s, \dots, t$ can be identified. As shown by Theorem 9 in the next section, it may become identifiable by adding a completely observed variable to it. Id_1 is used only as an auxiliary indicator to assert a conditionally identifiable $G_{1,2}$ to be identifiable. For example, when $G_{1,2}$ is subgraph (i) or (k), it is unconditional YR-identifiable since $Y_1 \perp\!\!\!\perp R_1$ and so $P(y_1)$ is identifiable.

The stepwise algorithm for checking identifiability consists of the following steps: (in the format of a Pascal program)

Input: A DAG model G.

Output: $Id_1, \dots, Id_T$.

1. *Initialization and checking the first time point*. Set $t = 1$.
   If $G_{1,1} \in \Omega_0$ then set $Id_1 = YR$ else set $Id_1 = U$.
2. *Checking identifiability one by one time point*. Repeat the following steps until $t = T - 1$:
   (a) When $G_{t,t+1} \in \Omega_1$, set $Id_{t+1} = YR$.

(b) When $G_{t,t+1} \in \Omega_2$,
if the missing process is not drop-out at time point $t+1$
then set $Id_{t+1} = YR$
else if $Id_t = YR$ or Y
then set $Id_{t+1} = YR$
else $Id_{t+1} = C$.

(c) When $G_{t,t+1} \in \Omega_3$,
if $Id_t = YR$ or Y then set $Id_{t+1} = YR$ else set $Id_{t+1} = C$.

(d) When $G_{t,t+1} \in \Omega_4$,
if $G_{t+1,t+2} \in \Omega_1$
then { if the missing process is drop-out at $t+1$ or $G_{t-1,t} \in \Omega_1 \cup \Omega_2 \cup \Omega_3$
then set $Id_{t+1} = YR$
else set $Id_{t+1} = Y$}
else set $Id_{t+1} = U$.

(e) When $G_{t,t+1} \in \Omega_5$,
if $G_{t+1,t+2} \in \Omega_1$ then set $Id_{t+1} = YR$ else set $Id_{t+1} = U$.

(f) When $G_{t,t+1} \notin \Omega_1 \cup \cdots \cup \Omega_5$, set $Id_{t+1} = U$.

For all cases, set $t = t+1$.

3. *Finally for every time point with conditional identifiability, check its adjacent time points.* For $Id_t = C$ for a time point t, then set $Id_t = YR$ if Id_{t-1} or $Id_{t+1} = YR$ or Y.

Theorem 8. *If the stepwise algorithm asserts that every subgraph in a sequence of subgraphs $G_{s,s+1}$, $G_{s+1,s+2}$, …, $G_{t-1,t}$, where $s \leqslant t$ is YR- (or Y- or conditionally YR-) identifiable, then $G_{s,t}$ is YR- (or Y- or conditionally YR-) identifiable by $D_{1,T}$.*

Example. Consider the DAG model in Fig. 2. The stepwise approach is shown in Table 1. From the last row of Table 1, we obtain that $G_{1,3}$ is YR-identifiable, $G_{3,4}$ cannot be asserted, $G_{4,5}$ is Y-identifiable and $G_{5,6}$ is YR-identifiable.

5. Adding completely observed variables to improve identifiability

In this section, we propose adding variables to an unidentifiable model such that it becomes identifiable. Let X denote a completely observed binary variable. We first

Table 1
The process of checking identifiability

t	$G_{t,t+1}$	Id_1	Id_2	Id_3	Id_4	Id_5	Id_6
Init.		U					
1	Ω_3	U	C				
2	Ω_1	U	C	YR			
3	Ω_5	U	C	YR	U		
4	Ω_4	U	C	YR	U	Y	
5	Ω_1	U	C	YR	U	Y	YR
Final		U	YR	YR	U	Y	YR

consider the subgraph of Fig. 3(b) under which the response indicator R_1 depends on the value of Y_1. The distribution of Y_1 and R_1 is

$$P(y_1, r_1) = P(y_1)P(r_1 \mid y_1).$$

We cannot identify $P(y_1)$ at this time point since there is no information about $P(R_1 = 0 \mid y_1)$ in observed data. To identify $P(y_1)$, we introduce a binary variable X such that $R_1 \perp\!\!\!\perp X \mid Y_1$, as shown in Fig. 3(c), where an undirected edge between X and Y_1 denotes an arrow which may be either from X to Y_1 or from Y_1 to X. Thus the conditional distribution of Y_1 and R_1 given $X = x$ can be written as

$$P(y_1, r_1 \mid x) = P(y_1 \mid x)P(r_1 \mid y_1).$$

For all h and x, let

$$\theta_{h|x} = P(Y_1 = h, R_1 = 1 \mid X = x), \quad p_{h|x} = P(Y_1 = h \mid X = x),$$
$$\rho_h = P(R_1 = 1 \mid Y_1 = h).$$

$\theta_{h|x}$ can be identified directly from observed data, and $p_{h|x}$ and ρ_h are the parameters to be identified. Note that in this case, $\rho_0 \neq \rho_1$ since this model can be classified into Fig. 3(a) if $\rho_0 = \rho_1$.

Theorem 9. *Subgraph* (c) *in Fig.* 3 *is YRX-identifiable by observed frequencies of* $(R_1 = 1, y_1, x)$ *and* $(R_1 = 0, x)$ *if and only if* $X \not\!\perp\!\!\!\perp Y_1$.

Theorem 9 can be applied to any other time point t. For example, if all subgraphs $G_{t,t+1}$ of a DAG G have type (j) in Ω_3, then the DAG model is conditionally *YR* identifiable. In this case, we can add a variable X which is connected by a unique arrow with some Y_t. Then $P(y_t, r_t)$ is identifiable and so is this DAG model with X. Adding the arrow from X to Y_1 corresponds to no missing at the initial time $t = 0$ of a longitudinal study, that is, Y_0 is observed completely.

Next, we consider subgraphs (o) and (p), as shown in Theorem 7, whose identifiability is asserted by checking the structure of $G_{t+1,t+2}$. Subgraphs (o) and (p) describe that the response indicator R_{t+1} depends on both values of Y_t and Y_{t+1}. In order to identify models (o) and (p), we introduce a binary variable X such that $R_{t+1} \perp\!\!\!\perp X | (Y_t, Y_{t+1})$, as shown in Fig. 5 (o′) and (p′) where the undirected edges may be arrows in any direction except for the cases constructing a cyclic graph. Regardless of the directions of arrows among Y_t, Y_{t+1} and X, the conditional distribution of Y_{t+1} and R_{t+1} given y_t, r_t and x can be written as

$$P(y_{t+1}, r_{t+1} \mid y_t, r_t, x) = P(y_{t+1} \mid y_t, x)P(r_{t+1} \mid y_{t+1}, y_t).$$

Define

$$\theta_{j|hx} = P(Y_{t+1} = j, R_{t+1} = 1 \mid Y_t = h, R_t = 1, X = x),$$

$$p_{j|hx} = P(Y_{t+1} = j \mid Y_t = h, X = x)$$

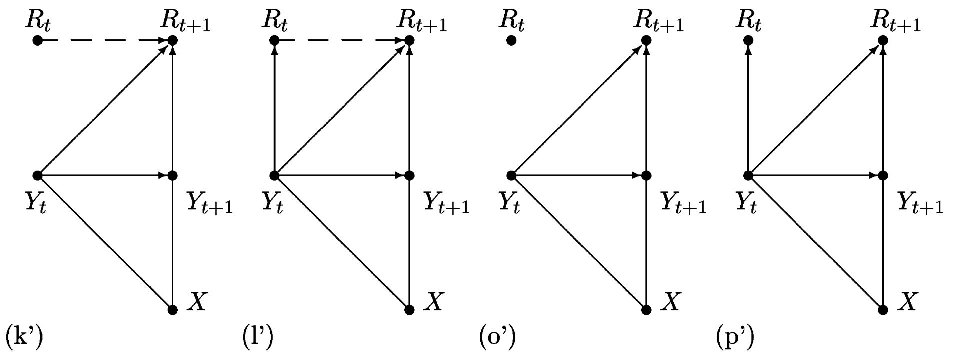

Fig. 5. Subgraphs with a variable X.

and

$$\rho_{hj} = P(R_{t+1} = 1 \mid Y_{t+1} = j, Y_t = h).$$

Parameters $\theta_{j|hx}$ can be identified directly from observed data.

Theorem 10. *Subgraphs* (o′) *and* (p′) *are conditionally YRX-identifiable* (*i.e.*, $P(y_t, r_t, y_{t+1}, r_{t+1}, x)$ *is identifiable if either* $P(y_t)$ *or* $P(y_{t+1})$ *is identifiable*) *by observed frequencies of* $(R_t = 1, R_{t+1} = 1, y_t, y_{t+1}, x)$, $(R_t = 1, R_{t+1} = 0, y_t, x)$, $(R_t = 0, R_{t+1} = 1, y_{t+1}, x)$ *and* $(R_t = 0, R_{t+1} = 0, x)$ *if* $X \not\perp\!\!\!\perp Y_{t+1} | Y_t$ *and* $Y_t \not\perp\!\!\!\perp Y_{t+1}$.

Lemma 3. *Under subgraphs* (o′) *and* (p′), $P(y_t)$ *is identifiable if* $Y_t \perp\!\!\!\perp R_t$ *or* $X \not\perp\!\!\!\perp Y_t$.

Subgraphs (k) and (l) of Fig. 4 describe that the response indicator R_{t+1} depends on values of Y_t, Y_{t+1} and R_t. There is no enough information on $P(R_{t+1} = 1 | R_t = 0, y_t, y_{t+1})$ in $D_{t,t+1}$, and thus it is not identifiable by $D_{t,t+1}$. When the missing process is drop-out, we have

$$P(R_{t+1} = 1 | R_t = 0, y_t, y_{t+1}) = 0,$$

for all t, y_t and y_{t+1}. Their identifiability becomes the same as that of models (o) and (p). Thus, we can add a variable X to subgraphs (k) and (l) such that they are identifiable, as shown in Fig. 5(k′) and (l′), where the dashed arrow from R_t to R_{t+1} denotes drop-out, that is, $R_{t+1} = 0$ if $R_t = 0$.

6. Summary

For binary outcomes in a longitudinal study, DAG models are used to describe nonignorable nonresponse mechanisms with the first-order Markov dependence. In this paper, we showed conditions for identifiability of the mechanisms, and proposed a stepwise approach for checking identifiability. DAG models can be used qualitatively to describe conditional independence among variables. The approach proposed in this paper asserts identification based on DAG structures and thus it can be applied to a study design.

For an unidentifiable model, a common approach is to impose some additional constraints on the parameters such that it becomes identifiable. For an unidentifiable model, we propose to add completely observed variables in the study design. They may promote the identifiability of the model.

Acknowledgments

We thank the editor and two referees for their helpful comments. This research was supported by NSFC, RFDP, NEYSF and Intel China Ltd.

Appendix A

Proof of Lemma 1. If $P(y_{t+1})$ and $p_{j|h}$ are identifiable and $Y_t \not\perp\!\!\!\perp Y_{t+1}$, then we can identify $P(y_t)$ by solving the equation

$$P(Y_{t+1}=1) = p_{1|0}P(Y_t=0) + p_{1|1}[1 - P(Y_t=0)].$$

Then $P(y_t, R_t = 0)$ can be identified by $P(y_t) - P(y_t|R_t=1)P(R_t=1)$. Thus the parameters λ_{hi} are identifiable. $\square$

Proof of Theorem 1. We first consider subgraphs (a) and (b). They satisfy $Y_{t+1} \perp\!\!\!\perp R_{t+1}$ and $\rho_{hij} = P(R_{t+1}=1) = \rho$ for all h, i and j. Thus $P(y_{t+1})$ and $p_{j|h}$ can be identified by $P(y_{t+1}|R_{t+1}=1)$ and $\theta_{j|h}/\rho$, respectively.

For the sufficiency, if $Y_t \perp\!\!\!\perp R_t$, then we can identify λ_{hi} by $P(Y_t=h|R_t=1)P(R_t=i)$. If $Y_t \not\perp\!\!\!\perp Y_{t+1}$, Lemma 1 then implies that λ_{hi} is identifiable. Hence $P(y_t, r_t, y_{t+1}, r_{t+1})$ is identifiable.

For the necessity, suppose that both $Y_{t+1} \perp\!\!\!\perp Y_t$ and $R_t \not\perp\!\!\!\perp Y_t$. Then we have $P(Y_t, R_t, Y_{t+1}, R_{t+1}) = P(Y_t, R_t)P(Y_{t+1}, R_{t+1})$, and thus data $D_{t+1,t+1}$are not useful for identifying $P(Y_t, R_t)$. Since $R_t \not\perp\!\!\!\perp Y_t$, similar to Fig. 3(b), we have that λ_{hi} is not identified by $D_{t,t}$ and so it is neither identified by $D_{t,t+1}$. Thus the identifiability of subgraphs (a) and (b) by $D_{t,t+1}$ implies $Y_{t+1} \not\perp\!\!\!\perp Y_t$ or $R_t \perp\!\!\!\perp Y_t$.

Next we consider subgraphs (c) and (d). Under these models, $\rho_{hij} = P(R_{t+1} = 1|Y_{t+1}=j) = \rho_j$ for all h, i and j. We have that $\rho_0 \neq \rho_1$ since subgraphs (c) and (d) with $\rho_0 = \rho_1$ become subgraphs (a) and (b), respectively. That $Y_{t+1} \not\perp\!\!\!\perp Y_t$ means

$$\frac{P(Y_{t+1}=0 \mid Y_t=0)}{P(Y_{t+1}=1 \mid Y_t=0)} \neq \frac{P(Y_{t+1}=0 \mid Y_t=1)}{P(Y_{t+1}=1 \mid Y_t=1)}.$$

Multiplying both sides by $P(R_{t+1}=1 \mid Y_{t+1}=0)/P(R_{t+1}=1 \mid Y_{t+1}=1)$, we have

$$\frac{P(R_{t+1}=1, Y_{t+1}=0 \mid Y_t=0)}{P(R_{t+1}=1, Y_{t+1}=1 \mid Y_t=0)} \neq \frac{P(R_{t+1}=1, Y_{t+1}=0 \mid Y_t=1)}{P(R_{t+1}=1, Y_{t+1}=1 \mid Y_t=1)}.$$

The above inequality is equivalent to

$$\begin{vmatrix} \theta_{0|0} & \theta_{1|0} \\ \theta_{0|1} & \theta_{1|1} \end{vmatrix} \neq 0.$$

The parameters ρ_0 and ρ_1 are identified because they are the solutions to the equation

$$\begin{pmatrix} \theta_{0|0} & \theta_{1|0} \\ \theta_{0|1} & \theta_{1|1} \end{pmatrix} \begin{pmatrix} 1/\rho_0 \\ 1/\rho_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

After ρ_0 and ρ_1 are identified, we can obtain $p_{j|h} = \theta_{j|h}/\rho_j$ and

$$P(Y_{t+1} = j) = P(R_{t+1} = 1, Y_{t+1} = j)/\rho_j.$$

From Lemma 1, we can identify λ_{hi}.

For necessity, similar to the proof of (a) and (b), if $Y_{t+1} \perp\!\!\!\perp Y_t$, then the subgraph $G_{t,t+1}$ becomes two separate parts, one of which for time $t+1$ is the same as Fig. 3(b). Thus we know that $P(Y_{t+1} = j, R_{t+1} = k)$ is not identifiable.

Finally we consider subgraphs (e) and (f). Under these models, $\rho_{hij} = P(R_{t+1} = 1 | Y_t = h) = \rho_h$ for all h, i and j. ρ_h and $p_{j|h}$ can be identified by $P(R_{t+1} = 1 | Y_t = h, R_t = 1)$ and $P(Y_{t+1} = j | Y_t = h, R_{t+1} = 1, R_t = 1)$, respectively. We show below that λ_{hi} is identifiable. Under subgraphs (e) and (f), we have

$$P(R_t = 1, y_t \mid R_{t+1} = k) = P(y_t \mid R_{t+1} = k) P(R_t = 1 \mid y_t).$$

Define

$$\phi_{h|k} = P(R_t = 1, Y_t = h \mid R_{t+1} = k),$$

$$\gamma_{h|k} = P(Y_t = h \mid R_{t+1} = k)$$

and

$$\tau_h = P(R_t = 1 \mid Y_t = h).$$

If $Y_t \perp\!\!\!\perp R_t$, then λ_{hi} can be identified by $P(Y_t = h | R_t = 1) P(R_t = i)$. Otherwise, we have $\tau_0 \neq \tau_1$. Note that if $Y_t \perp\!\!\!\perp R_{t+1}$, then subgraphs (e) and (f) become (a) and (b), respectively. Thus we have $Y_t \not\perp\!\!\!\perp R_{t+1}$ for (e) and (f), and then

$$\frac{P(Y_t = 0 \mid R_{t+1} = 0)}{P(Y_t = 1 \mid R_{t+1} = 0)} \neq \frac{P(Y_t = 0 \mid R_{t+1} = 1)}{P(Y_t = 1 \mid R_{t+1} = 1)}.$$

Multiplying both sides by $P(R_t = 1 | Y_t = 0)/P(R_t = 1 | Y_t = 1)$, the above inequality can be rewritten as

$$\frac{P(R_t = 1, Y_t = 0 \mid R_{t+1} = 0)}{P(R_t = 1, Y_t = 1 \mid R_{t+1} = 0)} \neq \frac{P(R_t = 1, Y_t = 0 \mid R_{t+1} = 1)}{P(R_t = 1, Y_t = 1 \mid R_{t+1} = 1)}.$$

Thus the following equation

$$\begin{pmatrix} \phi_{0|0} & \phi_{1|0} \\ \phi_{0|1} & \phi_{1|1} \end{pmatrix} \begin{pmatrix} 1/\tau_0 \\ 1/\tau_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

has a unique solution for τ_h. Thus we can identify $P(Y_t = h)$ by $P(R_t = 1, Y_t = h)/\tau_h$, and then identify λ_{hi} by $P(R_t = i | Y_t = h)P(Y_t = h)$. $\square$

Proof of Theorem 2. Under subgraphs (g) and (h), $\rho_{hij} = P(R_{t+1} = 1 | R_t = i) = \rho_i$ and thus can be identified by $D_{t,t+1}$ directly. Since $(R_t, R_{t+1}) \perp\!\!\!\perp Y_{t+1} | Y_t$, $p_{j|h} = P(Y_{t+1} = j | Y_t = h, R_{t+1} = 1, R_t = 1)$ can be identified by $D_{t,t+1}$.

We first consider the sufficiency. If $Y_{t+1} \not\perp\!\!\!\perp Y_t$, Lemma 1 then implies that λ_{hi} is identifiable if one of $P(y_t)$ or $P(y_{t+1})$ is identifiable. Hence subgraphs (g) and (h) are conditionally YR-identifiable by $D_{t,t+1}$ for both cases of $P(R_{t+1} = 1 | R_t = 0) = 0$ and > 0.

If $R_t \perp\!\!\!\perp Y_t$, then $\lambda_{hi} = P(Y_t = h | R_t = 1)P(R_t = i)$ is identifiable from $D_{t,t+1}$ directly. Hence subgraphs (g) and (h) are YR-identifiable.

When $P(R_{t+1} = 1 | R_t = 0) > 0$ but $R_t \not\perp\!\!\!\perp Y_t$, we show below that $P(y_{t+1})$ is identifiable, and thus subgraphs (g) and (h) are YR-identifiable. $P(R_t = 1, R_{t+1} = 1, Y_{t+1} = j)$ and $P(R_t = 0, R_{t+1} = 1, Y_{t+1} = j)$ are identifiable from $D_{t,t+1}$ directly. We have that

$$P(R_t = 1, R_{t+1} = 0, Y_{t+1} = j) = \lambda_{01} p_{j|0}(1 - \rho_1) + \lambda_{11} p_{j|1}(1 - \rho_1)$$

and

$$\begin{aligned} &P(R_t = 0, R_{t+1} = 0, Y_{t+1} = j) \\ &\quad = \frac{P(R_t = 0, R_{t+1} = 1, Y_{t+1} = j)}{P(R_{t+1} = 1 | R_t = 0)} P(R_{t+1} = 0 | R_t = 0) \end{aligned}$$

are identifiable. Hence

$$P(Y_{t+1} = j) = \sum_{i,k} P(R_t = i, R_{t+1} = k, Y_{t+1} = j)$$

is identifiable. Thus, when $P(R_{t+1} = 1 | R_t = 0) > 0$, we showed that λ_{hi} is identifiable.

For the necessity of the first result, suppose that $Y_{t+1} \perp\!\!\!\perp Y_t$ and $R_t \not\perp\!\!\!\perp Y_t$. Similar to the proof of Theorem 1, we have that λ_{hi} is not identifiable by $D_{t,t+1}$. $\square$

Proof of Theorem 3. Under subgraphs (i) and (j), $\rho_{hij} = P(R_{t+1} = 1 | R_t = i, Y_{t+1} = j) = \rho_{ij}$ for all h, i and j. That $Y_{t+1} \not\perp\!\!\!\perp Y_t$ is equivalent to

$$\frac{P(Y_{t+1} = 0 \mid Y_t = 0)}{P(Y_{t+1} = 0 \mid Y_t = 1)} \neq \frac{P(Y_{t+1} = 1 \mid Y_t = 0)}{P(Y_{t+1} = 1 \mid Y_t = 1)}.$$

Multiplying both sides by ρ_{10}/ρ_{11}, we have

$$\frac{P(R_{t+1} = 1, Y_{t+1} = 0 \mid Y_t = 0, R_t = 1)}{P(R_{t+1} = 1, Y_{t+1} = 0 \mid Y_t = 1, R_t = 1)} \neq \frac{P(R_{t+1} = 1, Y_{t+1} = 1 \mid Y_t = 0, R_t = 1)}{P(R_{t+1} = 1, Y_{t+1} = 1 \mid Y_t = 1, R_t = 1)},$$

which is equivalent to

$$\begin{vmatrix} \theta_{0|0} & \theta_{1|0} \\ \theta_{0|1} & \theta_{1|1} \end{vmatrix} \neq 0.$$

This implies that the equation

$$\begin{pmatrix} \theta_{0|0} & \theta_{1|0} \\ \theta_{0|1} & \theta_{1|1} \end{pmatrix} \begin{pmatrix} 1/\rho_{10} \\ 1/\rho_{11} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

has a unique solution for ρ_{10} and ρ_{11}. $p_{j|h}$ can be identified by $\theta_{j|h}/\rho_{1j}$. From Lemma 1, we have that λ_{hi} is identifiable if $P(y_t)$ or $P(y_{t+1})$ is identifiable.

Now we show that ρ_{0j} are identifiable. In the case of drop-out, we have $\rho_{0j} = 0$. Otherwise we can identify ρ_{0j} as follows. Rewrite ρ_{0j} as

$$\rho_{0j} = P(R_{t+1} = 1|R_t = 0,\, Y_{t+1} = j) = \frac{P(R_{t+1} = 1,\, Y_{t+1} = j|R_t = 0)}{P(Y_{t+1} = j|R_t = 0)}.$$

The numerator can be identified directly. The denominator can be expressed as

$$\begin{aligned} P(Y_{t+1} = j|R_t = 0) = {} & P(Y_{t+1} = j|\, Y_t = 0, R_t = 0)P(Y_t = 0|R_t = 0) \\ & + P(Y_{t+1} = j|\, Y_t = 1, R_t = 0)P(Y_t = 1|R_t = 0). \end{aligned}$$

For subgraphs (i) and (j), we have $P(Y_{t+1} = j|\, Y_t = h, R_t = 0) = P(Y_{t+1} = j|\, Y_t = h) = p_{j|h}$, and we have shown above that $p_{j|h}$ is identifiable. Also we have shown that λ_{hi} is identifiable, and hence so is $P(Y_t = i|R_t = 0)$. The above equation implies that the denominator $P(Y_{t+1} = j|R_t = 0)$ is identifiable, and thus ρ_{0j} can be identified. Therefore subgraphs (i) and (j) are identifiable.

For subgraphs (k) and (l), $\rho_{hij} = P(R_{t+1} = 1|\, Y_t = h, R_t = i) = \rho_{hi}$ for all h, i and j. $p_{j|h}$ can be identified by $\theta_{j|h}/P(R_{t+1} = 1|\, Y_t = h, R_t = 1)$. Thus, by Lemma 1, we deduce that λ_{hi} are identifiable by $D_{t,t+1}$ if $P(Y_t)$ or $P(Y_{t+1})$ is identifiable and $Y_t \not\perp\!\!\!\perp Y_{t+1}$. When $P(Y_t)$ or $P(Y_{t+1})$ is identifiable, λ_{hi} is identifiable if $Y_{t+1} \not\perp\!\!\!\perp Y_t$.

Below we show that ρ_{hi} is identifiable. In the case of drop-out, we have that $\rho_{h0} = 0$. In the other case, when λ_{hi} is identifiable, we have

$$\begin{pmatrix} \lambda_{00}p_{0|0} & \lambda_{10}p_{0|1} \\ \lambda_{00}p_{1|0} & \lambda_{10}p_{1|1} \end{pmatrix} \begin{pmatrix} 1/\rho_{00} \\ 1/\rho_{10} \end{pmatrix} = \begin{pmatrix} P(R_{t+1} = 1, R_t = 0,\, Y_{t+1} = 0) \\ P(R_{t+1} = 1, R_t = 0,\, Y_{t+1} = 1) \end{pmatrix}.$$

Both the matrix and the vector on the right-hand side are identifiable. If $Y_{t+1} \not\perp\!\!\!\perp Y_t$, then the matrix is non-singular, and thus ρ_{00} and ρ_{10} have a unique solution. $\rho_{h1} = P(R_{t+1} = 1|\, Y_t = h, R_t = 1)$ can be identified directly from data. Therefore subgraphs (k) and (l) are identifiable. □

Proof of Lemma 2. Define

$$\eta_{j|hz} = P(Y_{t+1} = j, R_{t+1} = 1 \mid Y_t = h, R_t = 1, Z_{t+2} = z),$$

$$\pi_{j|hz} = P(Y_{t+1} = j \mid Y_t = h, R_t = 1, Z_{t+2} = z), \quad \tau_{hj} = \rho_{h1j}.$$

From the Markov property $Z_{t+2} \perp\!\!\!\perp (Y_t, R_t) \mid (Y_{t+1}, R_{t+1})$ and the supposition $R_{t+1} \perp\!\!\!\perp Z_{t+2} \mid Y_{t+1}$, we have

$$P(Z_{t+2} = z|r_{t+1}, y_{t+1}, y_t, r_t) = P(Z_{t+2} = z|r_{t+1}, y_{t+1}) = P(Z_{t+2} = z|y_{t+1}),$$

that is, $Z_{t+2} \perp\!\!\!\perp (R_{t+1}, R_t, Y_t) | Y_{t+1}$. Thus we obtain $Z_{t+2} \perp\!\!\!\perp R_{t+1} | (Y_{t+1}, R_t, Y_t)$. So we have

$$\eta_{j|hz} = \pi_{j|hz}\tau_{hj}.$$

That $Z_{t+2} \not\!\perp\!\!\!\perp Y_{t+1}$ is equivalent to

$$\frac{P(Y_{t+1} = 0 | Z_{t+2} = 0)}{P(Y_{t+1} = 1 | Z_{t+2} = 0)} \neq \frac{P(Y_{t+1} = 0 | Z_{t+2} = 1)}{P(Y_{t+1} = 1 | Z_{t+2} = 1)}.$$

Multiplying both sides by $P(Y_t = h, R_t = 1 \mid Y_{t+1} = 0)/P(Y_t = h, R_t = 1 \mid Y_{t+1} = 1)$, we get

$$\frac{P(Y_{t+1} = 0, R_t = 1, Y_t = h \mid Z_{t+2} = 0)}{P(Y_{t+1} = 1, R_t = 1, Y_t = h \mid Z_{t+2} = 0)} \neq \frac{P(Y_{t+1} = 0, R_t = 1, Y_t = h \mid Z_{t+2} = 1)}{P(Y_{t+1} = 1, R_t = 1, Y_t = h \mid Z_{t+2} = 1)}.$$

This inequality is equivalent to

$$\frac{P(Y_{t+1} = 0 \mid R_t = 1, Y_t = h, Z_{t+2} = 0)}{P(Y_{t+1} = 1 \mid R_t = 1, Y_t = h, Z_{t+2} = 0)} \neq \frac{P(Y_{t+1} = 0 \mid R_t = 1, Y_t = h, Z_{t+2} = 1)}{P(Y_{t+1} = 1 \mid R_t = 1, Y_t = h, Z_{t+2} = 1)}.$$

Multiplying both sides by $P(R_{t+1} = 1 \mid R_t = 1, Y_t = h, Y_{t+1} = 0)/P(R_{t+1} = 1 \mid R_t = 1, Y_t = h, Y_{t+1} = 1)$, we get

$$\frac{\eta_{0|h0}}{\eta_{1|h0}} \neq \frac{\eta_{0|h1}}{\eta_{1|h1}}.$$

From the supposition that $\eta_{j|hz}$ is identifiable, we obtain a unique solution for τ_{hj} from the following equation:

$$\begin{pmatrix} \eta_{0|00} & \eta_{1|00} & 0 & 0 \\ \eta_{0|01} & \eta_{1|01} & 0 & 0 \\ 0 & 0 & \eta_{0|10} & \eta_{1|10} \\ 0 & 0 & \eta_{0|11} & \eta_{1|11} \end{pmatrix} \begin{pmatrix} 1/\tau_{00} \\ 1/\tau_{01} \\ 1/\tau_{10} \\ 1/\tau_{11} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

Since $R_t \perp\!\!\!\perp Y_{t+1} \mid Y_t$, we can identify $p_{j|h}$ by

$$p_{j|h} = P(Y_{t+1} = j \mid Y_t = h, R_t = 1) = \frac{P(R_{t+1} = 1, y_{t+1} \mid y_t, R_t = 1)}{\tau_{hj}}. \qquad \square$$

Proof of Theorem 4. Let $\tau_h = P(R_t = 1 | Y_t = h)$. If $\tau_0 = \tau_1$, then $Y_t \perp\!\!\!\perp R_t$, and thus $P(Y_t = h, R_t = i)$ can be identified by $P(Y_t = h | R_t = 1)P(R_t = i)$. Below we consider the case of $\tau_0 \neq \tau_1$. Since $Y_t \not\!\perp\!\!\!\perp R_{t+1}$, then we have

$$\frac{P(Y_t = 0 \mid R_{t+1} = 0)}{P(Y_t = 1 \mid R_{t+1} = 0)} \neq \frac{P(Y_t = 0 \mid R_{t+1} = 1)}{P(Y_t = 1 \mid R_{t+1} = 1)}.$$

Since $R_{t+1} \perp\!\!\!\perp R_t | Y_t$ under subgraphs (o) and (p), multiplying both sides by $P(R_t = 1 | Y_t = 0)/P(R_t = 1 | Y_t = 1)$, the above inequality can be rewritten as

$$\frac{P(R_t = 1, Y_t = 0 \mid R_{t+1} = 0)}{P(R_t = 1, Y_t = 1 \mid R_{t+1} = 0)} \neq \frac{P(R_t = 1, Y_t = 0 \mid R_{t+1} = 1)}{P(R_t = 1, Y_t = 1 \mid R_{t+1} = 1)}.$$

Thus the following equation

$$\begin{pmatrix} \phi_{0|0} & \phi_{1|0} \\ \phi_{0|1} & \phi_{1|1} \end{pmatrix} \begin{pmatrix} 1/\tau_0 \\ 1/\tau_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

has a unique solution for τ_h. Thus we can identify $P(Y_t = h)$ by $P(R_t = 1, Y_t = h)/\tau_h$, and then identify $P(Y_t = h, R_t = i)$ by $P(R_t = i | Y_t = h)P(Y_t = h)$. $\square$

Proof of Theorem 5. We first show that $p_{j|h}$ is identifiable and finally show that $P(y_t, y_{t+1})$ is identifiable by $p_{j|h}$.

Under subgraphs (m) and (n), we have that $R_t \perp\!\!\!\perp Y_{t+1} | Y_t$. We first consider the case that $G_{t+1,t+2} \in \Omega_1$ and $Y_{t+2} \not\perp\!\!\!\perp Y_{t+1}$. For $G_{t,t+1} \in \Omega_4$, it is impossible that $G_{t+1,t+2}$ is (a), (c) or (e). If $G_{t+1,t+2}$ is (f) and $R_{t+2} \not\perp\!\!\!\perp Y_{t+1}$, then treat R_{t+2} as Z_{t+2} of Lemma 2. Thus we have that $R_{t+1} \perp\!\!\!\perp R_{t+2} \mid Y_{t+1}$ and that $P(y_{t+1}, R_{t+1} = 1 \mid y_t, R_t = 1, r_{t+2})$ is identifiable. By Lemma 2, both $p_{j|h}$ and ρ_{h1j} are identifiable. If $G_{t+1,t+2}$ is (f) but $R_{t+2} \perp\!\!\!\perp Y_{t+1}$, then we can see (f) as (b). If $G_{t+1,t+2}$ is (b) or (d), then we have $R_{t+2} \perp\!\!\!\perp (Y_t, R_t, Y_{t+1}, R_{t+1}) | Y_{t+2}$, and thus $P(y_{t+1}, R_{t+1} = 1 | y_t, R_t = 1, y_{t+2})$ can be identified by $P(y_{t+1}, R_{t+1} = 1 | y_t, R_t = 1, y_{t+2}, R_{t+2} = 1)$. Treat Y_{t+2} as Z_{t+2} of Lemma 2. We have that $R_{t+1} \perp\!\!\!\perp Y_{t+2} | Y_{t+1}$ and $Y_{t+2} \not\perp\!\!\!\perp Y_{t+1}$. By Lemma 2, both $p_{j|h}$ and ρ_{h1j} are identifiable.

Next we consider the case that $G_{t+1,t+2} \in \Omega_5$ and $R_{t+2} \not\perp\!\!\!\perp Y_{t+1}$. In this case, treat R_{t+2} as Z_{t+2} of Lemma 2. Thus we have that $R_{t+1} \perp\!\!\!\perp R_{t+2} \mid Y_{t+1}$ and that $P(y_{t+1}, R_{t+1} = 1 \mid y_t, R_t = 1, r_{t+2})$ is identifiable. By Lemma 2, both $p_{j|h}$ and ρ_{h1j} are identifiable.

Finally, we show that λ_{hi} can be identified by using $p_{j|h}$. If $G_{t+1,t+2} \in \Omega_1$ and $Y_{t+2} \not\perp\!\!\!\perp Y_{t+1}$, then Theorem 1 implies that $P(y_{t+1})$ can be identified. If $G_{t+1,t+2} \in \Omega_5$ and $R_{t+2} \not\perp\!\!\!\perp Y_{t+1}$, Theorem 4 implies that $P(y_{t+1})$ can be identified by $D_{t+1,t+2}$. We obtain from Lemma 1 that λ_{hi} is identifiable, and thus $P(y_t, y_{t+1}) = p_{j|h} \sum_i \lambda_{hi}$ is identifiable, that is, $G_{t,t+1}$ is Y-identifiable by $D_{t,t+2}$. $\square$

Proof of Theorem 6. If conditions (i) and (ii) of Theorem 5 hold, then we have from the proof of Theorem 5 that $p_{j|h}$, ρ_{h1j} and λ_{hi} are identifiable. Thus we only need to show that ρ_{h0j} is identifiable. If the missing process is drop-out at time points t and $t+1$, then $\rho_{h0j} = 0$, and thus it is identifiable.

If the missing process is not drop-out, let

$$\omega_{h|k} = P(R_t = 0, Y_t = h | R_{t-1} = 1, Y_{t-1} = k)$$

and

$$\phi_{j|k} = P(R_{t+1} = 1, Y_{t+1} = j, R_t = 0 | R_{t-1} = 1, Y_{t-1} = k).$$

Since $G_{t-1,t}$ is a subgraph of Ω_1, Ω_2 or Ω_3, $Y_{t-1} \not\perp\!\!\!\perp Y_t$, and $G_{t,t+1}$ is Y-identifiable, $G_{t-1,t}$ is YR-identifiable and therefore so is $\omega_{h|k}$. $\phi_{j|k}$ is identifiable directly from data. If $Y_{t-1} \not\perp\!\!\!\perp Y_t$, we have

$$\frac{P(Y_t = 0 | Y_{t-1} = 0)}{P(Y_t = 0 | Y_{t-1} = 1)} \neq \frac{P(Y_t = 1 | Y_{t-1} = 0)}{P(Y_t = 1 | Y_{t-1} = 1)}.$$

The above inequality is equivalent to

$$\frac{P(R_t = 0, Y_t = 0 | R_{t-1} = 1, Y_{t-1} = 0)}{P(R_t = 0, Y_t = 0 | R_{t-1} = 1, Y_{t-1} = 1)} \neq \frac{P(R_t = 0, Y_t = 1 | R_{t-1} = 1, Y_{t-1} = 0)}{P(R_t = 0, Y_t = 1 | R_{t-1} = 1, Y_{t-1} = 1)}$$

for each subgraph in Ω_1, Ω_2 and Ω_3. For a DAG $G_{t,t+1}$, we have

$$P(r_t, y_t, r_{t+1}, y_{t+1}) = P(r_t, y_t) P(y_{t+1} | y_t) P(r_{t+1} | r_t, y_t, y_{t+1}),$$

and thus

$$P(y_{t+1} | y_t) = \frac{P(r_{t+1}, y_{t+1} | r_t, y_t)}{P(r_{t+1} | r_t, y_t, y_{t+1})}.$$

We obtain

$$\begin{aligned} \omega_{h|k} p_{j|h} \rho_{h0j} &= P(R_t = 0, Y_t = h | R_{t-1} = 1, Y_{t-1} = k) P(R_{t+1} = 1, Y_{t+1} = j | R_t = 0, Y_t = h) \\ &= P(R_t = 0, Y_t = h, R_{t+1} = 1, Y_{t+1} = j | R_{t-1} = 1, Y_{t-1} = k). \end{aligned}$$

Therefore the equation

$$\begin{pmatrix} \omega_{0|0} p_{j|0} & \omega_{1|0} p_{j|1} \\ \omega_{0|1} p_{j|0} & \omega_{1|1} p_{j|1} \end{pmatrix} \begin{pmatrix} \rho_{00j} \\ \rho_{10j} \end{pmatrix} = \begin{pmatrix} \phi_{j|0} \\ \phi_{j|1} \end{pmatrix}$$

has a unique solution for ρ_{h0j}. We showed that the subgraphs (m) and (n) are YR-identifiable. $\square$

Proof of Theorem 7. Under the subgraphs (o) and (p), $\rho_{hij} = P(R_{t+1} = 1 | Y_t = h, Y_{t+1} = j) = \rho_{hj}$ for all h, i and j, and $R_t \perp\!\!\!\perp Y_{t+1} | Y_t$. We first discuss the identifiability of $p_{j|h}$ and ρ_{hj} and then discuss the identifiability of λ_{hi}.

As in the proof of Theorem 5, we consider first the case where $G_{t+1,t+2}$ is in Ω_1. If $G_{t+1,t+2}$ is (f) and $R_{t+2} \not\perp\!\!\!\perp Y_{t+1}$, then we treat R_{t+2} as Z_{t+2} of Lemma 2. Thus $R_{t+1} \perp\!\!\!\perp R_{t+2} | Y_{t+1}$, and $P(y_{t+1}, R_{t+1} = 1 | y_t, R_t = 1, r_{t+2})$ is identifiable directly from data. By Lemma 2, both $p_{j|h}$ and $\rho_{h1j} = \rho_{hj}$ are identifiable. If $G_{t+1,t+2}$ is (f) but $R_{t+2} \perp\!\!\!\perp Y_{t+1}$, then we can see $G_{t+1,t+2}$ as (b). If $G_{t+1,t+2}$ is (b) or (d) and $Y_{t+2} \not\perp\!\!\!\perp Y_{t+1}$, then treat Y_{t+2} as Z_{t+2} of Lemma 2. Thus we have that $R_{t+1} \perp\!\!\!\perp Y_{t+2} \mid Y_{t+1}$ and that $P(y_{t+1}, R_{t+1} = 1 \mid y_t, R_t = 1, y_{t+2})$ can be identified by $P(y_{t+1}, R_{t+1} = 1 \mid y_t, R_t = 1, y_{t+2}, R_{t+2} = 1)$ since $R_{t+2} \perp\!\!\!\perp (Y_t, R_t, Y_{t+1}, R_{t+1}) | Y_{t+2}$ for (b) and (d). By Lemma 2, we obtain that both $p_{j|h}$ and $\rho_{h1j} = \rho_{hj}$ are identifiable.

Next we consider the case that $G_{t+1,t+2} \in \Omega_5$ and $R_{t+2} \not\perp\!\!\!\perp Y_{t+1}$. In this case, treat R_{t+2} as Z_{t+2} of Lemma 2. Thus we have that $R_{t+1} \perp\!\!\!\perp R_{t+2} \mid Y_{t+1}$ and that $P(y_{t+1}, R_{t+1} = 1 \mid y_t, R_t = 1, r_{t+2})$ is identifiable directly from data. By Lemma 2, we obtain that both $p_{j|h}$ and $\rho_{h1j} = \rho_{hj}$ are identifiable.

Finally, Theorems 1 and 4 imply that $P(y_{t+1})$ can be identified if ($G_{t+1,t+2} \in \Omega_1$ and $Y_{t+2} \not\perp\!\!\!\perp Y_{t+1}$) or ($G_{t+1,t+2} \in \Omega_5$ and $R_{t+2} \not\perp\!\!\!\perp Y_{t+1}$). Then, according to Lemma 1, λ_{hi} is identifiable when $Y_{t+1} \not\perp\!\!\!\perp Y_t$. Thus we showed that $G_{t,t+1}$ is YR-identifiable by $D_{t,t+2}$. $\square$

Proof of Theorem 8. First, it is obvious that identifiability by $D_{t,t+1}$ implies identifiability by $D_{1,T}$. Next, since the G has the first-order Markov dependence, the joint distributions $P(y_1, r_1, \ldots, y_T, r_T)$ and $P(y_1, \ldots, y_T)$ can be factorized respectively as

$$P(y_1, r_1, \ldots, y_T, r_T) = P(y_1, r_1)P(y_2, r_2|y_1, r_1)\ldots P(y_T, r_T|y_{T-1}, r_{T-1})$$

and

$$P(y_1, \ldots, y_T) = P(y_1)P(y_2|y_1)\ldots P(y_T|y_{T-1}).$$

Thus if $G_{i,i+1}$ is identifiable by $D_{i,i+1}$, $D_{i,i+2}$ or $D_{i-1,i+2}$ for $i = s, s+1, \ldots, t-1$, then $G_{s,t}$ is identifiable by $D_{1,T}$.

Step 1 of the algorithm can be justified directly from Figs. 3(a) and (b). Step 2 can be justified by the theorems in Section 3. Step 3 revises a conditional YR-identifiability to unconditional YR-identifiability if its adjacent time point is YR- or Y-identifiable. $\square$

Proof of Theorem 9. Fig. 3(c) is a special case of Figs. 4(b) and (c), where X corresponds to Y_{t+1} in (b) and Y_t in (c). The result follows immediately from Theorem 1. $\square$

Proof of Theorem 10. The conditional dependence $X \not\perp\!\!\!\perp Y_{t+1} | Y_t$ means

$$\frac{P(Y_{t+1} = 0 \mid Y_t = h, X = 0)}{P(Y_{t+1} = 1 \mid Y_t = h, X = 0)} \neq \frac{P(Y_{t+1} = 0 \mid Y_t = h, X = 1)}{P(Y_{t+1} = 1 \mid Y_t = h, X = 1)}.$$

Multiplying both sides by $P(R_{t+1} = 1 \mid Y_{t+1} = 0, Y_t = h)/P(R_{t+1} = 1 \mid Y_{t+1} = 1, Y_t = h)$, we get

$$\frac{\theta_{0|h0}}{\theta_{1|h0}} \neq \frac{\theta_{0|h1}}{\theta_{1|h1}},$$

since $R_{t+1} \perp\!\!\!\perp X | (Y_t, Y_{t+1})$. Thus the equation

$$\begin{pmatrix} \theta_{0|00} & \theta_{1|00} & 0 & 0 \\ \theta_{0|01} & \theta_{1|01} & 0 & 0 \\ 0 & 0 & \theta_{0|10} & \theta_{1|10} \\ 0 & 0 & \theta_{0|11} & \theta_{1|11} \end{pmatrix} \begin{pmatrix} 1/\rho_{00} \\ 1/\rho_{01} \\ 1/\rho_{10} \\ 1/\rho_{11} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

has a unique solution for ρ_{hj}. Then $p_{j|hx}$ can be identified by $\theta_{j|hx}/\rho_{hj}$. Hence $P(y_{t+1}, r_{t+1}|y_t, r_t, x)$ is identifiable.

We shall show below that $P(y_t, r_t, x)$ is conditionally identifiable. $P(y_t, R_t = 1)$ can be identified by $P(y_t|R_t = 1)P(R_t = 1)$. If $P(y_t)$ is identifiable, then $P(x|y_t, r_t)$ can be identified by $P(x|y_t, R_t = 1)$ since $X \perp\!\!\!\perp R_t | Y_t$, and then $P(y_t, R_t = 0)$ by $P(y_t) - P(y_t, R_t = 1)$. Hence $P(y_t, r_t, x)$ is identifiable if $P(y_t)$ is identifiable.

If $P(y_{t+1})$ is identifiable, we can identify $P(y_t)$ by solving

$$
\begin{aligned}
P(Y_{t+1} = 1) = {} & P(Y_{t+1} = 1 | Y_t = 0)P(Y_t = 0) \\
& + P(Y_{t+1} = 1 | Y_t = 1)[1 - P(Y_t = 0)]
\end{aligned}
$$

for $Y_t \not\perp\!\!\!\perp Y_{t+1}$ where

$$
\begin{aligned}
P(Y_{t+1} = j | Y_t = h) &= \frac{P(Y_{t+1} = j, R_{t+1} = 1 | Y_t = h)}{P(R_{t+1} = 1 | Y_t = h, Y_{t+1} = j)} \\
&= \frac{P(Y_{t+1} = j, R_{t+1} = 1 | Y_t = h, R_t = 1)}{\rho_{hj}}.
\end{aligned}
$$

Thus, from the above result, we showed that $P(y_t, r_t, x)$ is identifiable. $\square$

Proof of Lemma 3. The condition $Y_t \perp\!\!\!\perp R_t$ is obvious. If $X \not\perp\!\!\!\perp Y_t$, we can identify $P(y_t)$ by solving

$$
P(x) = P(Y_t = 1)P(x| Y_t = 1) + P(Y_t = 0)P(x| Y_t = 0),
$$

where $P(x|y_t)$ is identified by $P(x|y_t, R_t = 1)$ since $X \perp\!\!\!\perp R_t | Y_t$. $\square$

References

[1] R.L. Chambers, A.H. Welsh, Log-linear models for survey data with non-ignorable non-response, J. Roy. Statist. Soc. Ser. B 55 (1993) 157–170.

[2] D.R. Cox, N. Wermuth, Multivariate Dependencies: Models, Analysis and Interpretation, Chapman & Hall, London, 1996.

[3] P.J. Diggle, M.G. Kenward, Informative drop-out in longitudinal data analysis (with discussion), Appl. Statist. 43 (1994) 49–93.

[4] R.E. Fay, Causal models for patterns of nonresponse, J. Amer. Statist. Assoc. 81 (1986) 354–365.

[5] G.M. Fitzmaurice, N.M. Laird, G.E.P. Zahner, Multivariate logistic models for incomplete binary responses, J. Amer. Statist. Assoc. 91 (1996) 99–108.

[6] G.F.V. Glonek, On identifiability in models for incomplete binary data, Statist. Probab. Lett. 41 (1999) 191–197.

[7] S.L. Lauritzen, Graphical Models, Oxford University Press, Oxford, 1996.

[8] R.J.A. Little, Nonresponse adjustments in longitudinal surveys: models for categorical data, Bull. Internat. Statist. Inst. 15 (1) (1985) 1–15.

[9] R.J.A. Little, Pattern-mixture models for multivariate incomplete data, J. Amer. Statist. Assoc. 88 (1993) 125–134.

[10] R.J.A. Little, D.B. Rubin, Statistical Analysis with Missing Data, Wiley, New York, 1987.
[11] J. Pearl, Probabilistic Reasoning in Intelligent Systems, Morgan Kaufmann, San Mateo, 1988.
[12] T.J. Rothenberg, Identification in parametric models, Econometrica 39 (1971) 577–591.
[13] A. Rotnitzky, J.M. Robins, D.O. Scharfstein, Semiparametric regression for repeated outcomes with nonignorable nonresponse, J. Amer. Statist. Assoc. 93 (1998) 1321–1339.